OPEN ACCESS

University of BRISTOL

Voice, A., Tresadern, G., van Vlijmen, H., & Mulholland, A. J. (2019). Limitations of ligand-only approaches for predicting the reactivity of covalent inhibitors. *Journal of Chemical Information and Modeling*. https://doi.org/10.1021/acs.jcim.9b00404

Peer reviewed version

Link to published version (if available):
10.1021/acs.jcim.9b00404

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via American Chemical Society at https://pubs.acs.org/doi/10.1021/acs.jcim.9b00404 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

Limitations of ligand-only approaches for predicting the reactivity of covalent inhibitors

Angus Voice [a], Gary Tresadern [b], Herman van Vlijmen [b], Adrian Mulholland [a]

[a] Centre of Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom. [b] Computational Chemistry, Janssen Research & Development, Janssen Pharmaceutica N. V., Turnhoutseweg 30, B-2340 Beerse, Belgium

KEYWORDS Covalent, TCI, Reactivity prediction, Drug discovery, Proton affinity, Density Functional Theory.

Abstract: Covalent inhibition has undergone a resurgence and is an important modern-day drug design and chemical biology approach. To avoid off-target interactions, and to fine tune reactivity, the ability to accurately predict reactivity is vitally important for the design and development of safer and more effective covalent drugs. Several ligand-only metrics have been proposed that promise quick and simple ways of determining covalent reactivity. In particular, we examine proton affinity and reaction energies calculated with the density functional B3LYP-D3/6-311+G**//B3LYP-D3/6-31G* method to assess the reactivity of a series of α,β-unsaturated carbonyl compounds that form covalent adducts with cysteine. We demonstrate that, whilst these metrics correlate well with experiment for a diverse range of covalent fragments, these approaches fail for predicting the reactivity of drug-like compounds. We conclude that ligand-only metrics such as proton affinity and reaction energies do not capture determinants of reactivity in situ and fail to account for important factors such as conformation, solvation and intramolecular interactions.

Predicting covalent reactivity is an important goal in targeted covalent inhibitor (TCI) design. In contrast to conventional reversible inhibitors, TCIs form a covalent attachment with their target, resulting in several advantageous properties. These include increased potency, selectivity, and residence time all leading to distinct pharmacodynamic properties.[1] The half-life of the covalent complex (which may be long, compared with the clearance of the free TCI and synthesis rate of the target protein) can offer pharmacokinetic advantages. Efficacy can be extended beyond what would be expected due to the half-life of the free drug alone in plasma. There are now several recently marketed covalent kinase inhibitors, for example (Figure 1).[2] However, toxicity concerns are an important consideration when pursuing a covalent mechanism of action in drug design, owing to the intrinsic reactivity of TCIs.[3] In addition, clearance and cross-reactivity/selectivity can be problematic.[4] A typical approach is the addition of a reactive covalent warhead onto a potent reversible inhibitor of the target, in the expectation that selectivity will be improved by the optimized fit of the reversible molecule for the particular protein target.[5] Covalent warheads can target various amino acids including cysteine, serine, tyrosine and lysine.[6] Amongst these, acrylamides are often used to target cysteine residues; acrylamide reactivity can be modulated via the amine and β-carbon substituents of the motif.[7]

As the field of TCIs has blossomed over recent years, some model/simplified experimental and computational approaches have been proposed to assist in predicting optimal reactivity. Experimental approaches have been used to estimate the general reactivity of covalent fragments with sulfur-containing species representing cysteine, typically glutathione (GSH).[8] This is often reported as GSH $t_{1/2}$, the half-life of the reaction of a covalent compound with glutathione. The GSH trapping assessment is readily accessible within drug discovery due to its legacy as a model for reactive toxicity, particularly of metabolites.[9] Therefore, such data is relatively easy to generate compared to protein binding reactivity or kinetics and serves as a convenient metric for comparing warhead reactivity with cysteine in TCI design.[10]

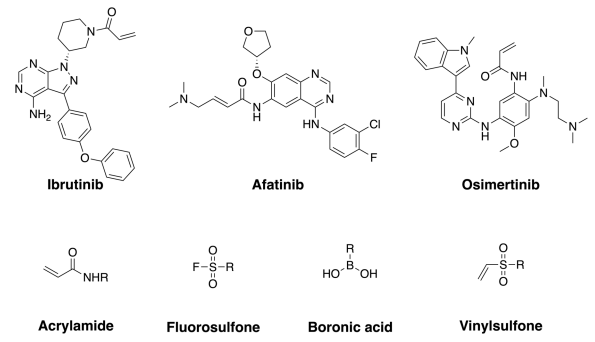


Figure 1. Recently approved acrylamide containing covalent kinase inhibitors[2] and alternative covalent warheads. Ibrutinib targets BTK and is used as a treatment for multiple B cell cancers. Afatinib and osimertinib both target EGFR and are treatments for non-small cell lung carcinoma.

With regard to computational chemistry methods, modeling the reaction in the protein-ligand complex with QM/MM methods has shown success[11–15] but remains complex and relatively time-consuming for drug discovery. Instead, one approach has

been to reduce the covalent reaction between acrylamides and their protein target to a simple ligand-only reactivity metric.

Ligand-only methods have shown some success in predicting reactivity trends for small covalent fragments.[16–18] Examples of these methods include using calculated proton affinity (PA), reaction energies (ΔE_{react}), kinetic barrier heights ($\Delta G^{\ddagger}$), electrophilicity index and NMR chemical shifts to predict covalent reactivity. Recent work from Houk *et al*. elaborated the full reaction mechanism for a small set of α,β-unsaturated carbonyl compounds illustrating that both kinetics and thermodynamics are important determinants of covalent reactivity of Michael acceptors with methyl thiolate.[19] Complementary to this, Lonsdale *et al* used density functional theory (DFT) to calculate reaction energies of a diverse set of covalent fragments. Whilst they set out calculating some QM transition state (TS) barriers, most of their work focused on calculating the energy difference between reactants and adduct equivalent to species **1** and **5** (ΔE_{react} in Figure 2). They found these to correlate well with GSH reactivity (R^2=0.69).[17] Furthermore, Flanagan *et al* found that calculated reaction barrier heights ($\Delta G^{\ddagger}$) correlate strongly with GSH reactivity (R^2=0.92).[20] $\Delta G^{\ddagger}$ is the energy difference between species **1** and the transition state in Figure 2, and reflects the overall reactivity rate (GSH $t_{1/2}$) of the reaction. PA provides an estimation of the free energy difference between the covalent-thiol adduct and its corresponding conjugate base (**4** and **5** respectively, ΔG_{PA} in Figure 2). Krishnan et al found that PA values correlate well with β-elimination rates of covalent fragments from sulfur containing species, which vary based on the acidity of the α-carbon adjacent to the carbonyl group of the acrylamide.[16] More recently, the electrophilicity index, derived from QM calculations has been shown to correlate well with experimental data. However, the authors note the difficulty in reactivity prediction using this method for larger compounds and non-terminal acrylamides.[21]

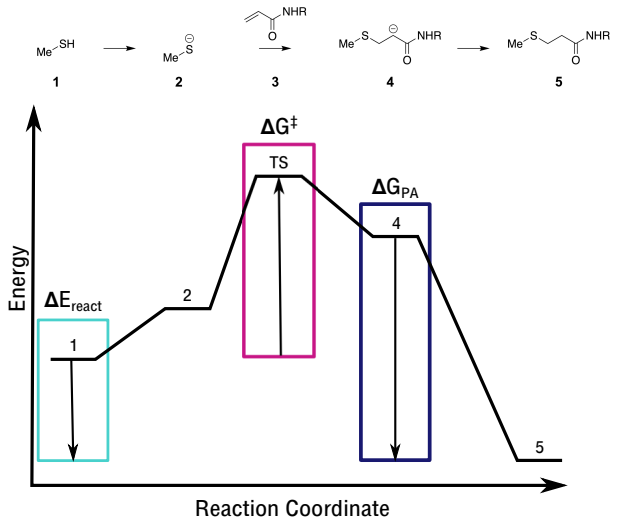


Figure 2. A schematic view of thiol addition to a generic α,β-unsaturated carbonyl compound. The energy profile highlights the chemical species commonly used in ligand-only reactivity metrics: reaction energy (ΔE_{react}), reaction barrier height ($\Delta G^{\ddagger}$) and proton affinity (ΔG_{PA})

Proton affinities and reaction energies (ΔE_{react}) are particularly attractive reactivity metrics, as their computational prediction is relatively straightforward from quantum mechanical (QM) calculations.[22] PA and ΔE_{react} are reactivity metrics that provide insight into the thermodynamic stability of the reaction. Although the TS is not explicitly used, PA and ΔE_{react} calculations give a good approximation to overall experimental reactivity as the enolate intermediate **4** that results from thio-Michael addition between cysteine and an acrylamide (Figure 2) is close in energy and geometry to the TS.[17] When considering an appropriate computational protocol to predict reactivity for drug molecules, the ability to consistently and reliably generate results for various input molecules is crucial because such an approach needs to be applicable to libraries of hundreds of analogues. The automated QM calculation of transition states presents a challenge because of the difficulty to reach geometrical convergence for saddle point structures whereas calculating PA and ΔE_{react} is more straightforward. However, new tools for TS searches are emerging and have been applied to thio-Michael reactivity, although the authors note the difficulty due to the typically flat nature of the potential energy surface (PES) around the TS.[23] In addition, it is difficult to accurately represent solvation changes with continuum solvation models, particularly for reactions involving charged intermediates, although good results can be achieved with explicit treatment of solvent.[24]

Our aim here is to investigate if two common ligand-only reactivity metrics, PA and ΔE_{react}, can accurately and reliably be used to predict biologically relevant covalent reactivity. We use a large number of compounds ranging from small covalent fragments, to large drug-like molecules to demonstrate how the simple ligand-only reactivity trends based on PA and ΔE_{react} compare favorably with GSH reactivity for similar small fragments but break down for larger drug like molecules.

Methods

A previous study used QM calculations to calculate proton affinities with the B3LYP functional and the 6-311+G(d) basis set.[16] Here we use a modified protocol, appropriate for proton affinity calculations of larger compounds. We performed gas phase geometry optimizations at the B3LYP-D3/6-31G(d) level of theory, followed by single point energies in solution using B3LYP-D3/6-311+G(d,p) level of theory. This has been shown to be sufficient for the calculation of proton affinity values for simple organic molecules of this type.[16,25,26] The inclusion of dispersion corrections to B3LYP has been recommended in QM calculations[27], particularly when modelling cysteine reactivity.[28] Although B3LYP has known limitations in modelling reactivity[29–32], it provides an acceptable balance of accuracy and speed in this context for predicting relative reactivities. We have also tested the M06-2X functional, but found it made no appreciable difference to the results (Supporting Information, Section 4). We included solvation by using the Poisson-Boltzmann Finite (PBF) element implicit water solvation model using a probe radius of 1.4 Å,[33,34] which has been shown to reproduce experimental solvation energies for anionic species with good accuracy.[35] All QM calculations were performed with Jaguar v8.5-13.[36]

For larger, drug-like molecules, conformational sampling becomes important. An assessment of the variation of PA with respect to the conformation of each molecule was performed. Six representative molecules were chosen. A conformational search was performed on each using the Macromodel tool in Maestro v2018-3.[37] The mixed torsional/low-mode sampling approach was used with default settings, using water as a solvent. An energy window of 5 kcal/mol was applied to discard higher energy conformations. This search resulted in over 2100 conformations for the 6 molecules combined. Two distinct conformations per molecule were extracted: the lowest energy conformation; and the next lowest energy with a root mean square deviation (RMSD) > 2.0 Å with respect to the first. This RMSD cutoff is typically used to distinguish structurally different conformations of drug-like molecules. These conformations were submitted to QM minimization at the B3LYP-D3/6-311+G(d,p) level of theory again using the PBF solvation model. All outputs were submitted to frequency calculations to confirm stationary points as minima, and to obtain zero-point energy for thermal and entropic corrections. These two minima were then used as inputs for the PA calculations as described above. PAs were calculated according to Figure 2. This involved calculating the difference in energy of the two states, **4** and **5** to obtain ΔG_{PA}. These ΔG_{PA} values were then normalized relative to the most acidic compound within each data set, providing what we refer to as $\Delta\Delta G_{PA}$ (kcal mol^{-1}). Reaction energy (ΔE_{react}) values were calculated by taking the difference in energy between states **1** and **5**. Since we already had optimized structures for state **5**, state **1** was generated simply by breaking the S-C bond to minimize towards the local reactant energy minimum. Geometry optimization was performed at the B3LYP-D3/6-31G(d) level in the gas phase, followed by single point energy calculations in aqueous solution at the B3LYP-D3/6-311+G(d,p) level of theory.

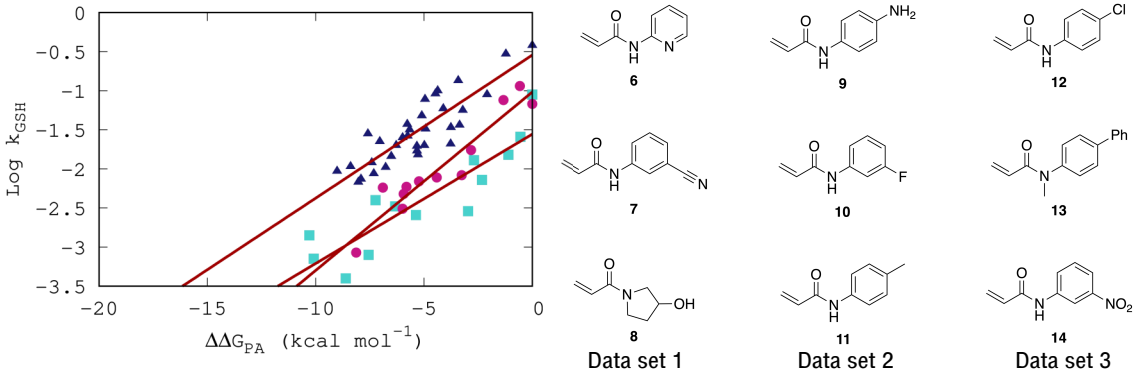


Figure 3. Plot of GSH reactivity vs relative proton affinity ($\Delta\Delta G_{PA}$) for data set 1,[20] R^2=0.79 turquoise squares, data set 2,[18] R^2=0.90 magenta circles, data set 3,[38] R^2=0.75 blue triangles. Representative structures from each data set are shown. There is a positive correlation observed between $\Delta\Delta G_{PA}$ and log k_{GSH} for each data set. More acidic compounds (represented by a more positive $\Delta\Delta G_{PA}$ value) have faster experimental reactivity rates.

Results

A literature search was performed to identify data sets that contained small molecule compounds bearing acrylamide functionality and associated GSH reactivity data. The fragments identified had a molecular weight range of 133 to 271 g mol^{-1}, and between 1 and 6 rotatable bonds. Using the method outlined above, we calculated proton affinity values for three data sets containing a total of 37 unique compounds. For each data set, a plot of the relative proton affinity ($\Delta\Delta G_{PA}$) against the GSH reactivity rate (log k) reveals a strong correlation: data set 1: R^2=0.79; data set 2: R^2=0.90; and data set 3: R^2=0.75 (Figure 3). A more negative $\Delta\Delta G_{PA}$ value corresponds to a more acidic compound and as expected these acidic compounds also have a faster GSH reactivity rate.

To test our approach for larger, more conformationally complex compounds, we performed a literature search for data sets that contained 'drug-like' compounds, bearing structural similarity to already approved covalent drugs on the market that also had published GSH reactivity data. The compounds had higher MW and increased flexibility, MW in the range 352 to 659 g mol^{-1}, and number of rotatable bonds from 6 to 13. We retrieved 3 data sets that contained 6, 11, and 9 compounds respectively. The first data set was based on a combination of anilinoquinoline and anilinoquinazoline EGFR inhibitors[18], the second on a series of bis-anilino-pyrimidine EGFR inhibitors[18] and the third were based on the chemical scaffold of pyrazolopyrimidine BTK inhibitors containing an inverted cyanoacrylamide warhead. [39]

For these larger molecules, conformational sampling could be important. Therefore, we performed a test for the conformational dependence of PA. We took the 6 compounds from data set 4, and identified the lowest energy conformation, and also the next lowest energy but structurally different conformation (>2 Å RMSD to the first, see methods) for each molecule. These conformations were then used to calculate the PA. The different conformations show significantly different PAs. Figure 4 shows there are large variations between the $\Delta\Delta G_{PA}$ calculated for the two alternative low energy conformations of each molecule. There are some examples such as compound **15** (Figure 4) where the PA for the two different conformations varies by as much as 9.3 kcal mol^{-1}, whereas others are closer. In addition, we calculated a Boltzmann weighted $\Delta\Delta G_{PA}$ value for all compounds in data set 4, averaged over 4 low energy conformers per molecule (Supporting Information, section 2). However, we find that there is no significant improvement in correlation between PA and experimental reactivity when using this approach (Figure S1, Supporting Information).

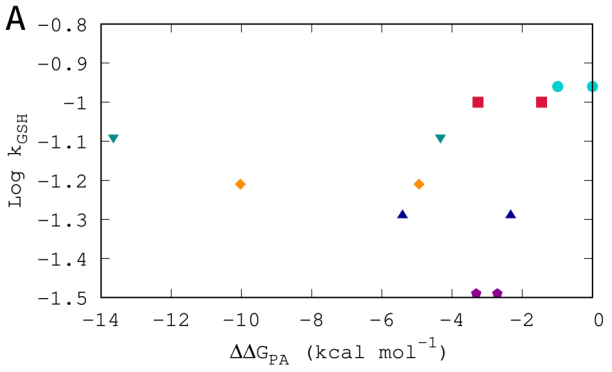


Figure 4. (A) Plot of $\Delta\Delta G_{PA}$ vs log(k_{GSH}) for the two conformers of each compound from data set 4, each pair of conformers for the same molecule is represented by different shape labels. (B) Structures of the lowest energy conformer and alternative low energy conformer with RMSD > 2 Å for compound **15**. Hydrogen atoms are omitted for clarity. These conformers exhibit a large variation in calculated $\Delta\Delta G_{PA}$, caused by a rotation of the acrylamide side chain.

The strong conformational dependence of the PA suggested that it would not be an ideal metric for large compounds. We investigated this by performing PA calculations using the method outlined above for all the compounds in these new data sets, but only using the lowest energy conformation. This involved a total of 25 unique drug-like compounds across the three data sets. For each data set, the plot of the relative proton affinity ($\Delta\Delta G_{PA}$) against the GSH reactivity rate (log k) revealed no correlation: data set 1: R^2=0.12; data set 2: R^2=0.00; and data set 3: R^2=0.14 (Figure 5).

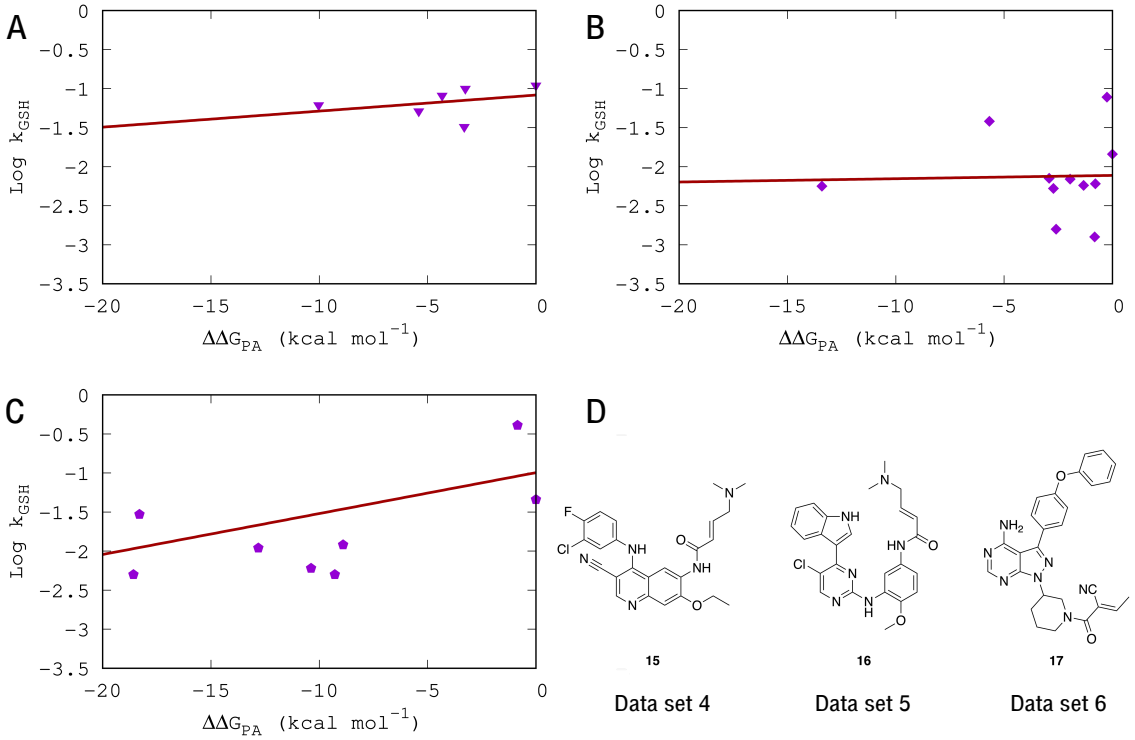


Figure 5. Plots of GSH reactivity vs relative proton affinity ($\Delta\Delta G_{PA}$) for (A) data set 4,[18] R^2=0.12, (B) data set 5,[18] R^2=0.00 (C) data set 6,[39] R^2=0.14. (D) Representative structures from each data set. There is no correlation observed between $\Delta\Delta G_{PA}$ and log k_{GSH} for each data set. More acidic compounds (represented by a more positive $\Delta\Delta G_{PA}$ value) would be expected to have faster experimental reactivity rates, but this is not observed by the large molecules in these sets.

As shown in Figure 3, the calculated PA for small fragments appears to be predictive for reactivity with GSH. Therefore, we examined it as a surrogate for predicting the reactivity of the larger molecules. We chose all the molecules from data set 4 and converted these to small fragments, so that only the core scaffold from the parent structure remained but each fragment was unique (Figure 6). These fragments are analogous to the types of compounds we studied in data sets 1-3. This approach is similar to the truncation algorithm used by Palazzesi et al[21], but we had to modify the approach slightly to avoid duplication of fragments. The calculated PAs for the lowest energy conformation of the full molecules and the fragments are shown in Table 1. The results suggest that the fragment calculations are less predictive for the reactivity of larger molecules and do not correlate with experiment. For instance, the most reactive molecule identified from experiment (F22) was predicted to be the

least reactive based on the fragment $\Delta\Delta G_{PA}$ calculation. Thus, even for the simpler case of GSH compared with protein reactivity, additional parameters beyond the local atomic and electronic environment are important.

Table 1. Calculated $\Delta\Delta G_{PA}$ values at the B3LYP-D3/6-311+G// B3LYP-D3/6-31G* level for fragments and the corresponding parent compound, and experimental reactivity ranking. The predicted relative rank order from calculations is shown in parentheses.**

Fragment	Fragment $\Delta\Delta G_{PA}$	'Parent' $\Delta\Delta G_{PA}$	Experimental ranking [a]
F15	–11.6 (4)	–4.3 (3)	4
F18	–7.0 (5)	0.0 (6)	6
F19	–15.2 (1)	–5.4 (2)	2
F20	–13.2 (3)	–3.3 (5)	5
F21	–13.3 (2)	–10.0 (1)	3
F22	0.0 (6)	–3.3 (4)	1

[a] Experimental ranking is based on most reactive being assigned ranking of 1, and least reactive a ranking of 6.

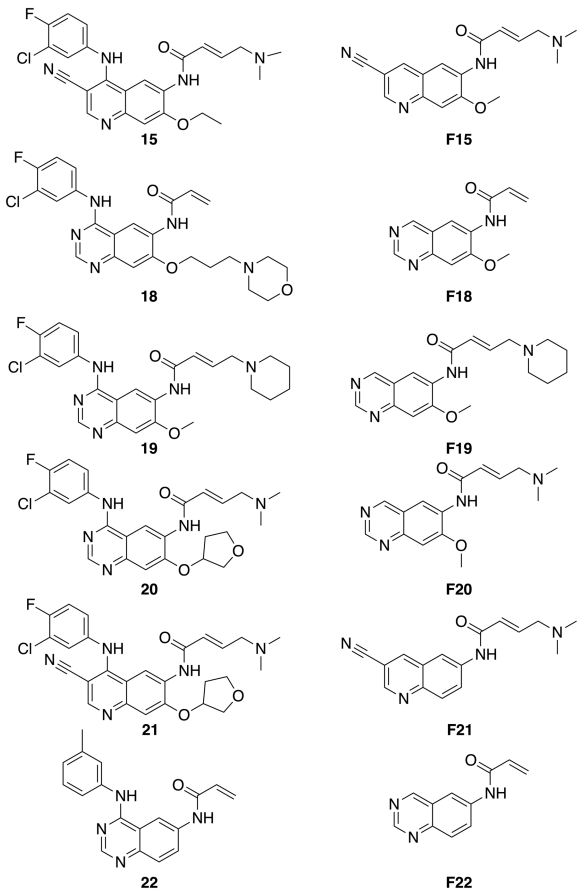


Figure 6. Fragments of drug-like compounds from data set 4 used here to calculate fragment $\Delta\Delta G_{PA}$ values.

We also investigated how calculated ΔE_{react} values, a commonly used alternative computational reactivity metric, correlate with the experimental reactivity. ΔE_{react} values were calculated for all 68 compounds from data sets 1-6 (see methods). Generally, a poor correlation between ΔE_{react} and experimental GSH reactivity was observed (Figure 7). Across all the data sets, $\Delta\Delta G_{PA}$ generally performed better as a reactivity metric than calculated ΔE_{react} values, although neither did as well as would be needed for a drug discovery lead optimization setting. In short, four of the six data sets of close analogues showed no appreciable correlation ($R^2 < 0.5$).

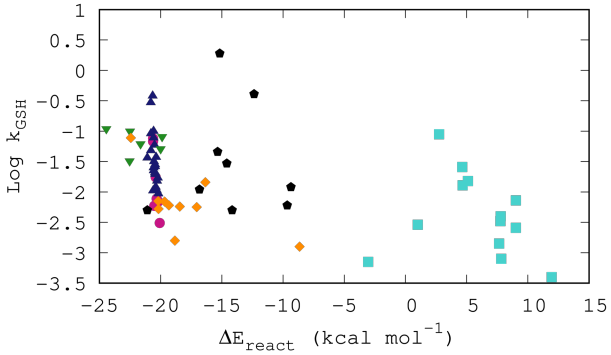


Figure 7. Plot of GSH reactivity vs reaction energy, ΔE_{react} for data set 1, R^2=0.05 turquoise squares; data set 2, R^2=0.68 magenta circles; data set 3, R^2=0.33 blue triangles; data set 4, R^2=0.09 green triangles; data set 5, R^2=0.50 orange diamonds; data set 6, R^2=0.00 black pentagons.

Discussion

The calculation of PA values for small covalent fragments bearing α,β-unsaturated functionality initially appears to be a promising covalent reactivity metric. For compounds of this type, we observe that increased acidity at the α-carbon of the acrylamide group correlates with faster GSH reactivity rates, as expected. This reflects the stability of the enolate intermediate (compound **4** in Figure 2) in the covalent thio-Michael reaction mechanism, which leads to faster reaction. The resulting plots show that $\Delta\Delta G_{PA}$ values correlate strongly with experimental GSH reactivity for small fragment compounds. The results are similar to previously published work[16] that found proton affinity to correlate with β-elimination rates with R^2=0.96.

For the larger and more complex compounds, no correlation is observed between $\Delta\Delta G_{PA}$ values and GSH reactivity. To understand this result, we investigated the effect of conformational differences in our protocol by examining in detail the 6 compounds in data set 4. Two sources of potential error were identified, the first being that different input conformations significantly impact the calculated PA for each compound. The difference in PA between conformer pairs obtained from the conformational search (see Methods) can be as large as 9.3 kcal mol^{-1} (compound **15**, data set 4, Figure 4A and 5). In this particular case, there was a rotation of the substituted acrylamide-thiol adduct leading to an additional NH-N interaction in one of the conformers that was not present in the other (Figure 4B). The second source of possible error resulted from the difference in conformation between the neutral and ionized species. After a geometry optimization, conformational changes in the ionized species were often observed, arising due to the change in hybridization (sp^3 to sp^2) of the alpha-carbon upon deprotonation causing a change in intramolecular interactions. Rotations of flexible groups attached to aromatic rings and movements of

the acrylamide-thiol adduct were also observed, suggesting the ionized species were optimized to alternate local minima from the neutral species. We used more stringent convergence criteria during our optimizations in order to address this, however, it did not significantly change the results (Figure S2, Supporting Information). These sources of error result in an overestimation of the calculated PA, suggesting an inaccurate comparison of the relative change in energy associated with proton affinity with the inclusion of conflicting conformational effects, and therefore explain the lack of correlation between PA and experimental reactivity.

An additional factor that could lead to errors in the calculation of $\Delta\Delta G_{PA}$ values is the contribution to the free energy from low energy vibrational modes. These low energy modes are particularly important for the large, flexible compounds in data set 4, 5 and 6, and are difficult to calculate accurately and can therefore lead to errors in the free energy calculation.[40] To investigate the affects that these modes have on our free energy values, we used the Quasi Harmonic Oscillator (QHO) approximation[41] to scale all frequencies under 100 cm^{-1} up to 100 cm^{-1} (Table S2, Supporting Information). These calculations were performed using GoodVibes, a python script from the Paton lab.[42] Whilst some small changes to the ΔG_{PA} values were observed, we conclude that these changes are small and do not significantly improve the correlation between $\Delta\Delta G_{PA}$ and experimental GSH reactivity, therefore offering no improvement to the predictive power of ligand-only methods for the larger molecules.

The ΔE_{react} calculations also showed generally poor correlation with GSH reactivity. This is in contrast to previous work by Lonsdale and colleagues who reported a correlation of R^2=0.69. Their study combined diverse molecules and warheads.[17] ΔE_{react} values may offer reasonable reactivity prediction for compounds that differ greatly in reactivity, but will generally not predict reactivity within a chemical series, as shown by our results. It is also likely that reaction energy calculations are also more susceptible to conformational effects than $\Delta\Delta G_{PA}$, due to the greater variability in geometry optimization of methyl thiolate relative to the reactive warhead.

Although previous studies have reported success in using a fragment based approach to improve correlation between ligand-only reactivity metrics and experimental reactivity,[21] we do not find it useful in the context of PA. Inspection of the experimental data and molecular structures casts further doubt on using a fragment approach to predict the reactivity of larger molecules. For instance, compounds **15** (data set 4), **23**, **24** and **25** from data set 5 (Figure 8) all contain essentially the same fragment (a dimethylamine substituted phenylacrylamide group), but their experimental reactivity differs by ~2 log units. This variation in reactivity highlights that modifications to compounds distal from the site of reaction can have a significant impact on reactivity (conformational effects will also contribute).

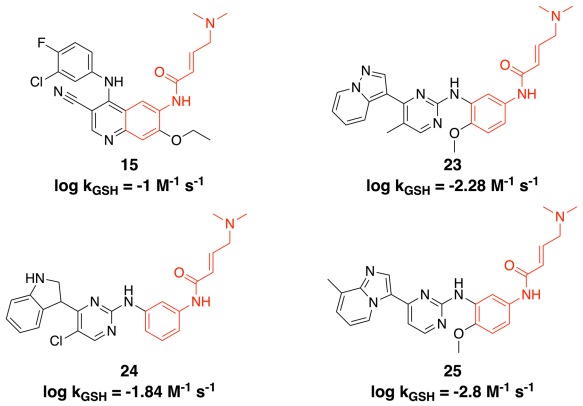


Figure 8. Four compounds that contain the same dimethylamine substituted phenylacrylamide fragment (highlighted in red) but have a wide range of experimental GSH reactivity, covering almost 2 log units.[18]

All other factors being equal, the change in reactivity caused by a small structural modification may be captured by PA calculations. Indeed, there could be scenarios in lead optimization where analogues are so similar that this method delivers results that translate into useful design prioritization. PA and reaction energies seem to be attractive reactivity metrics due to their simplicity. For PA, only geometry optimizations of the covalent thiol adduct and its conjugate base are required, negating the need to control the position of methyl thiolate during transition state searches and calculations of reaction energies. However, we see here that, for many common modifications in the medicinal chemistry data sets, further effects are at play. The calculation of PA, reaction energies and reaction barrier heights all require careful consideration of conformational variations in each compound as it is important to ensure that differences in the energy calculation do not arise from competing conformational effects. Lonsdale et al. found that the calculation of reaction energy barriers through transition state optimizations to be particularly challenging for large 'drug like' compounds and opted to calculate reaction energies instead.[17]

For biological reactivity, PA calculations require careful consideration that input geometries represent local minima corresponding to the noncovalently or covalently bound conformation, as opposed to global energy minima adopted in solution.[43] Reactivity of covalent fragments with GSH in solution is likely to be less constrained, but still dominated by a small number of conformations. Although conformational searches and low energy conformers were selected for our calculations, it is possible that these low energy conformers are not representative of the conformations that react with GSH in solution. Even Boltzmann weighted PA values, averaged over 4 low energy conformers that would be expected to dominate reactivity did not improve reactivity predictions (Supporting Information, Figure S1). In the protein environment it is easy to envisage that the reaction proceeds from the bound conformation of the ligand placed optimally for reactivity with the cysteine in the protein pocket. Replicating these reactive conformations is difficult to achieve for PA calculations and in part explains the lack of predictive ability of ligand-only approaches and indicates the need for treatment of biological reactivity in situ in biological targets.

Conclusions

In summary, calculated proton affinity values can serve as a useful reactivity metric for small molecule fragment compounds. However, we find proton affinity to be a poor reactivity metric for larger, 'drug-like' compounds. Calculated reaction energies are generally worse as a reactivity metric than proton affinities for both fragments and drug-like compounds. Our work shows that ligand-only approaches such as the calculation of reaction energies, barrier heights and proton affinity values are insufficient for reliable covalent reactivity prediction. These methods oversimplify covalent determinants of reactivity, and do not account for factors that will affect the reactivity of a covalent drug 'in-situ' (e.g. effects of the environment within a protein target). Calculated PA and reactivity are conformation dependent; experimental studies of ligand-only reactivity may also be confounded by conformational effects. More detailed methods are needed to predict reactivity of covalent drugs in a protein environment, which include the (e.g. electrostatic) effects of the environment, the bound conformation and solvation of the covalent inhibitor in the target site; such methods, capable of modeling reactivity within proteins, including combined QM/MM techniques.[44,45]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Molecular structures, calculated proton affinity values and reaction energies for all compounds studied (PDF)

AUTHOR INFORMATION

Corresponding Author

* E-mail: Adrian.Mulholland@bristol.ac.uk

Author Contributions

Angus Voice (A.V.), Gary Tresadern (G.T.), Herman van Vlijmen (H.v.V.) and Adrian Mulholland (A.M.) conceived the project, A.V. and G.T. performed experiments, A.V. and G.T. analyzed results, A.V. and G.T. wrote the manuscript, all authors contributed to the final manuscript.

ORCID

Angus Voice: 0000-0003-0640-1553
Gary Tresadern: 0000-0002-4801-1644
Herman van Vlijmen: 0000-0002-1915-3141
Adrian Mulholland: 0000-0003-1015-4567

ACKNOWLEDGMENTS

This work has been supported by Janssen through a PhD studentship to AV. AJM thanks EPSRC for funding to CCP-BioSim (www.ccpbiosim.ac.uk) [EP/M022609/1].

This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol – http://www.bristol.ac.uk/acrc/.

ABBREVIATIONS

TCI, targeted covalent inhibitor; PA, proton affinity; QM, quantum mechanical; GSH, glutathione; DFT, density functional theory; PES, potential energy surface; TS, transition state.

REFERENCES

(1) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.

(2) Ghosh, A. K.; Samanta, I.; Mondal, A.; Liu, W. R. Covalent Inhibition in Drug Discovery. *ChemMedChem* **2019**, *14*, 889–906.

(3) Johnson, D. S.; Weerapana, E.; Cravatt, B. F. Strategies for Discovering and Derisking Covalent, Irreversible Enzyme Inhibitors. *Future Med. Chem.* **2010**, *2*, 949–964.

(4) Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chemie Int. Ed.* **2016**, *55*, 13408–13421.

(5) Zhang, J.; Yang, P. L.; Gray, N. S. Targeting Cancer with Small Molecule Kinase Inhibitors. *Nat. Rev. Cancer* **2009**, *9*, 28–39.

(6) Zhao, Z.; Bourne, P. E. Progress with Covalent Small-Molecule Kinase Inhibitors. *Drug Discov. Today* **2018**, *23*, 727–735.

(7) Gehringer, M.; Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology.

(8) MacFaul, P. A.; Morley, A. D.; Crawford, J. J. A Simple in Vitro Assay for Assessing the Reactivity of Nitrile Containing Compounds. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 1136–1138.

(9) Baillie, T. A. Metabolism and Toxicity of Drugs. Two Decades of Progress in Industrial Drug Metabolism.

(10) Böhme, A.; Thaens, D.; Paschke, A.; Schüürmann, G. Kinetic Glutathione Chemoassay to Quantify Thiol Reactivity of Organic Electrophiless - Application to α,β-Unsaturated Ketones, Acrylates, and Propiolates. *Chem. Res. Toxicol.* **2009**, *22*, 742–750.

(11) Callegari, D.; Ranaghan, K. E.; Woods, C. J.; Minari, R.; Tiseo, M.; Mor, M.; Mulholland, A. J.; Lodola, A. L718Q Mutant EGFR Escapes Covalent Inhibition by Stabilizing a Non-Reactive Conformation of the Lung Cancer Drug Osimertinib. *Chem. Sci.* **2018**, *9*, 2740–2749.

(12) Capoferri, L.; Lodola, A.; Rivara, S.; Mor, M. Quantum Mechanics/Molecular Mechanics Modeling of Covalent Addition between EGFR–Cysteine 797 and *N* -(4-Anilinoquinazolin-6-Yl) Acrylamide. *J. Chem. Inf. Model.* **2015**, *55*, 589–599.

(13) Lodola, A.; Capoferri, L.; Rivara, S.; Tarzia, G.; Piomelli, D.; Mulholland, A.; Mor, M. Quantum Mechanics/Molecular Mechanics Modeling of Fatty Acid Amide Hydrolase Reactivation Distinguishes Substrate from Irreversible Covalent Inhibitors. *J. Med. Chem.* **2013**, *56*, 2500–2512.

(14) Schirmeister, T.; Kesselring, J.; Jung, S.; Schneider, T. H.; Weickert, A.; Becker, J.; Lee, W.; Bamberger, D.; Wich, P. R.; Distler, U.; Tenzer, S.; Johé, P.; Hellmich, U. A.; Engels, B. Quantum Chemical-Based Protocol for the Rational Design of Covalent Inhibitors. *J. Am. Chem. Soc.* **2016**, *138*, 8332–8335.

(15) Schmidt, T. C.; Welker, A.; Rieger, M.; Sahu, P. K.; Sotriffer, C. A.; Schirmeister, T.; Engels, B. Protocol for Rational Design of Covalently Interacting Inhibitors. *ChemPhysChem* **2014**, *15*, 3226–3235.

(16) Krishnan, S.; Miller, R. M.; Tian, B.; Mullins, R. D.; Jacobson, M. P.; Taunton, J. Design of Reversible, Cysteine-Targeted Michael Acceptors Guided by Kinetic and Computational Analysis. *J. Am. Chem. Soc.* **2014**, *136*, 12624–12630.

(17) Lonsdale, R.; Burgess, J.; Colclough, N.; Davies, N. L.; Lenz, E. M.; Orton, A. L.; Ward, R. A. Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity. *J. Chem. Inf. Model.* **2017**, *57*, 3124–3137.

(18) Ward, R. A.; Anderton, M. J.; Ashton, S.; Bethel, P. A.; Box, M.; Butterworth, S.; Colclough, N.; Chorley, C. G.; Chuaqui, C.; Cross, D. A. E.; Dakin, L. A.; Debreczeni, J. É.; Eberlein, C.; Finlay, M. R. V.; Hill, G. B.; Grist, M.; Klinowska, T. C. M.; Lane, C.; Martin, S.; Orme, J. P.; Smith, P.; Wang. F.; Waring, M. J. Structure- and Reactivity-Based Development of Covalent Inhibitors of the Activating and Gatekeeper Mutant Forms of the Epidermal Growth Factor Receptor (EGFR). *J. Med. Chem.* **2013**, *56*, 7025–7048.

(19) Krenske, E. H.; Petter, R. C.; Houk, K. N. Kinetics and Thermodynamics of Reversible Thiol Additions to Mono- and

Diactivated Michael Acceptors: Implications for the Design of Drugs That Bind Covalently to Cysteines. *J. Org. Chem.* **2016**, *81*, 11726–11733.

(20) Flanagan, M. E.; Abramite, J. A.; Anderson, D. P.; Aulabaugh, A.; Dahal, U. P.; Gilbert, A. M.; Li, C.; Montgomery, J.; Oppenheimer, S. R.; Ryder, T.; Schuff, B. P.; Uccello, D. P.; Walker, G. S.; Wu, Y.; Brown, M. F.; Chen, J. M.; Hayward, M. M.; Noe, M. C.; Obach, R. S.; Philippe, L; Shanmugasundaram, V; Shapiro, M. J.; Starr, J.; Stroh, J.; Che, Y. Chemical and Computational Methods for the Characterization of Covalent Reactive Groups for the Prospective Design of Irreversible Inhibitors. *J. Med. Chem.* **2014**, *57*, 10072–10079.

(21) Palazzesi, F.; Grundl, M. A.; Pautsch, A.; Weber, A.; Tautermann, C. S. A Fast Ab Initio Predictor Tool for Covalent Reactivity Estimation of Acrylamides. *J. Chem. Inf. Model.* **2019**, acs.jcim.9b00316.

(22) Kolboe, S. Proton Affinity Calculations with High Level Methods. *J. Chem, Theory. Comput*. **2014**, 10, 3123-3128

(23) Jacobson, L. D.; Bochevarov, A. D.; Watson, M. A.; Hughes, T. F.; Rinaldo, D.; Ehrlich, S.; Steinbrecher, T. B.; Vaitheeswaran, S.; Philipp, D. M.; Halls, M. D.; Friesner, R. A. Automated Transition State Search and Its Application to Diverse Types of Organic Reactions. *J. Chem. Theory Comput.* **2017**, *13*, 5780–5797.

(24) Liu, Z.; Patel, C.; Harvey, J. N.; Sunoj, R. B. Mechanism and Reactivity in the Morita–Baylis–Hillman Reaction: The Challenge of Accurate Computations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30647–30657.

(25) Thapa, B.; Schlegel, H. B. Density Functional Theory Calculation of pK_a's of Thiols in Aqueous Solution Using Explicit Water Molecules and the Polarizable Continuum Model. *J. Phys. Chem. A* **2016**, *120*, 5726–5735.

(26) Fu, Y.; Liu, L.; Li, R. Q.; Liu, R.; Guo, Q. X. First-Principle Predictions of Absolute pK_a's of Organic Acids in Dimethyl Sulfoxide Solution. *J. Am. Chem. Soc.* **2004**, *126*, 814–822.

(27) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. A Practical Guide to Modelling Enzyme-Catalysed Reactions. *Chem. Soc. Rev.* **2012**, *41*, 3025.

(28) Fekete, A.; Komá romi, I. Modeling the Archetype Cysteine Protease Reaction Using Dispersion Corrected Density Functional Methods in ONIOM-Type Hybrid QM/MM Calculations; the Proteolytic Reaction of Papain †. *Phys. Chem. Chem. Phys* **2016**, *18*, 32847.

(29) Smith, J. M.; Jami Alahmadi, Y.; Rowley, C. N. Range-Separated DFT Functionals Are Necessary to Model Thio-Michael Additions. *J. Chem. Theory Comput.* **2013**, *9*, 4860–4865.

(30) Awoonor-Williams, E.; Walsh, A. G.; Rowley, C. N. Modeling Covalent-Modifier Drugs. *Biochim. Biophys. Acta - Proteins Proteomics* **2017**, *1865*, 1664–1675.

(31) Bennie, S. J.; van der Kamp, M. W.; Pennifold, R. C. R.; Stella, M.; Manby, F. R.; Mulholland, A. J. A Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase. *J. Chem. Theory Comput.* **2016**, *12*, 2689–2697.

(32) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728.

(33) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Honig, B.; Ringnalda, M.; Goddard, W. A. Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory. *J. Am. Chem. Soc.* **1994**, *116*, 11875–11882.

(34) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. New Model for Calculation of Solvation Free Energies: Correction of Self-Consistent Reaction Field Continuum Dielectric Theory for Short-Range Hydrogen-Bonding Effects. *J. Phys. Chem.* **1996**, *100*, 11775–11788.

(35) Senn, H. M.; Ziegler, T.; Nielsen, P. A.; Jaroszewski, J. W.; Norrby, P.-O.; Liljefors, T.; Cramer, C. J.; Truhlar, D. G.; Orozco, M.; Luque, F. J.; Tomasi, J.; Mennucci, B.; Cammi, R.; Pliego, J. R.; Riveros, J. M. Selected Publications for Further Reading on Continuum Solvation Models. *J. Am. Chem. Soc* **1996**, *100*, 597.

(36) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142.

(37) Schrödinger Release 2018-3: Maestro, Schrödinger, LLC, New York, NY, 2018. Maestro | Schrödinger.

(38) Cee, V. J.; Volak, L. P.; Chen, Y.; Bartberger, M. D.; Tegley, C.; Arvedson, T.; McCarter, J.; Tasker, A. S.; Fotsch, C. Systematic Study of the Glutathione (GSH) Reactivity of *N* -Arylacrylamides: 1. Effects of Aryl Substitution. *J. Med. Chem.* **2015**, *58*, 9171–9178.

(39) Bradshaw, J. M.; McFarland, J. M.; Paavilainen, V. O.; Bisconte, A.; Tam, D.; Phan, V. T.; Romanov, S.; Finkle, D.; Shu, J.; Patel, V.; Ton, T.; Li, X.; Loughhead, D. G.; Nunn, P. A.; Karr, D. E.; Gerritsen, M. E.; Funk, J. O.; Owens, T. D.; Verner, E.; Brameld, K. A.; Hill, R. J.; Goldstein, D. M.; Taunton, J. Prolonged and Tunable Residence Time Using Reversible Covalent Kinase Inhibitors. *Nat. Chem. Biol.* **2015**, *11*, 525–531.

(40) Haworth, N. L.; Wang, Q.; Coote, M. L. Modeling Flexible Molecules in Solution: A PKa Case Study. *J. Phys. Chem. A* **2017**, *121*, 5217–5225.

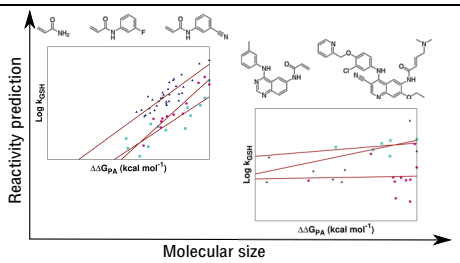
(41) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J. Phys. Chem. B* **2011**, *115*, 14556–14562.

(42) Funes-Ardoiz, I.; Paton, R. S. GoodVibes: GoodVibes v1.0.1. **2016**.

(43) Foloppe, N.; Chen, I.-J. Towards Understanding the Unbound State of Drug Compounds: Implications for the Intramolecular Reorganization Energy upon Binding. *Bioorg. Med. Chem.* **2016**, *24*, 2159–2189.

(44) Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W. Biomolecular Simulations: From Dynamics and Mechanisms to Computational Assays of Biological Activity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, *9*, e1393.

(45) Amaro, R. E.; Mulholland, A. J. Multiscale Methods in Drug Design Bridge Chemical and Biological Complexity in the Search for Cures. *Nat. Rev. Chem.* **2018**, *2*, 0148.



Insert Table of Contents artwork here